

PERAMALAN KASUS COVID-19 DI DKI JAKARTA DENGAN MODEL ARIMA

Nurul Qomariasih

Politeknik Siber dan Sandi Negara (Poltek SSN) Bogor Jawa Barat, Indonesia

Email : nurul.qomariasih@poltekssn.ac.id

INFO ARTIKEL

Diterima

1 Juni 2021

Direvisi

10 Juni 2021

Disetujui

20 Juni 2021

Keywords:

COVID-19; ARIMA; modelling;
forecasting

ABSTRACT

February 2020 was the first time that Indonesian citizens were diagnosed with Covid-19. Until now, the disease caused by the Corona Virus has not subsided, it has even been declared a Global Pandemic. The purpose of this study was to find a Time Series prediction model for the number of positive cases of Covid-19 in one of the cities with the largest number of infections in Indonesia, namely Jakarta. This study uses data from Open Data Jakarta with a time span of 1 September 2020 to 30 November 2020. The model used is the Auto Regressive Integrated Moving Average (ARIMA). The resulting ARIMA parameters are (0,1,1), (1,1,0), and (1,1,1), but the best model is ARIMA (1,1,1) where all parameters are significant with $p\text{-value} < \alpha$. The data used for training is 70% and 30% is used for testing. The resulting forecast value is very similar to the actual value. Thus, the model made is quite good in predicting positive cases of COVID-19 in Jakarta in the next month.

ABSTRAK

Februari 2020 merupakan waktu awal warga Indonesia didiagnosa positif Covid-19. Hingga kini, penyakit yang dikarenakan oleh Virus Corona ini belum mereda, bahkan dinyatakan sebagai Pandemi Global. Tujuan penelitian ini adalah mencari model prediksi *Time Series* untuk jumlah kasus positif Covid-19 di salah satu kota dengan jumlah infeksi terbesar di Indonesia yaitu Jakarta. Penelitian ini menggunakan data dari *Open Data Jakarta* dengan rentang waktu 1 September 2020 hingga 30 November 2020. Model yang digunakan adalah *Auto Regressive Integrated Moving Average* (ARIMA). Parameter ARIMA yang dihasilkan yaitu (0,1,1), (1,1,0), dan (1,1,1), namun model terbaiknya adalah ARIMA (1,1,1) dimana semua parameternya signifikan dengan $p\text{-value} < \alpha$. Data yang digunakan untuk *training* sebesar 70% dan yang digunakan untuk *testing* sebanyak 30%. Nilai peramalan yang dihasilkan sangat mirip dengan nilai aktualnya. Dengan demikian, model yang dibuat sudah cukup baik dalam meramalkan kasus positif COVID-19 di Jakarta dalam satu bulan kedepan.

Kata Kunci:

COVID-19; ARIMA; pemodelan;
peramalan

How to cite:

Qomariasih, Nurul (2021) Peramalan Kasus Covid-19 di DKI Jakarta dengan Model Arima. *Jurnal Syntax Transformation* 2(6). https://doi.org/10.46799/jurnal_syntax_transformation.v2i6.306

E-ISSN:

2721-2769

Published by:

Ridwan Institute

Pendahuluan

Coronavirus Disease (COVID-19) adalah virus RNA untai positif yang besar, berselubung, dan dapat dibagi menjadi 4: alfa, beta, delta, dan gamma; dimana alfa dan beta adalah jenis yang menginfeksi manusia (Nemmara & Thompson, 2018). Tidak hanya menginfeksi, tetapi penyakit ini juga menular. Penyakit ini menginfeksi sistem pernapasan, seseorang yang memiliki penyakit bawaan seperti penyakit jantung, diabetes dan kanker akan lebih tinggi resiko mengalami penyakit yang lebih serius akibat virus ini. Media penularan virus ini yakni melalui tetesan air liur atau keluarnya cairan dari hidung ketika seseorang yang terinfeksi bersin atau batuk. Belum diketahui secara jelas mengenai prevalensi penyakit ini, disebabkan saat ini prevalensi penyakit COVID-19 sangat dinamis (Niehus et al., 2020). Berdasarkan “*Our World in Data*” dan “*JHU CSSE COVID-19 Data*”, terdapat total kasus di seluruh dunia sebanyak 177 juta di seluruh dunia dan meninggal dunia sebanyak 3,84 juta jiwa. Terlebih lagi, terdapat beberapa kasus yang disebabkan oleh virus corona jenis baru. Di Indonesia, virus corona mulai menyerang warga pada tanggal 2 Maret 2020 dan masih menginfeksi hingga saat ini. Protokol kesehatan serta peraturan pembatasan sosial pun masih diberlakukan, sehingga virus corona diharapkan tidak menginfeksi lebih banyak penduduk di kemudian hari. Kondisi sarana kesehatan pun harus menjadi perhatian utama di saat pandemi global ini, banyak pasien tidak tertangani akibat melonjaknya jumlah kasus positif COVID-19. Sehingga pada penelitian ini, akan dilakukan peramalan jumlah kasus COVID-19 di salah satu kota dengan jumlah terinfeksi terbesar di Indonesia, yaitu DKI Jakarta. Dengan demikian, melalui prediksi statistik, dapat membantu sistem penanganan dalam memberikan

pelayanan kesehatan. Serta penilaian awal tingkat keparahan infeksi dan penularan dapat membantu mengukur pandemi potensi COVID-19 dan mengantisipasi kemungkinan jumlah kematian pada akhir pandemi (Jung et al., 2020). Model prediksi yang digunakan adalah *Auto Regressive Integrated Moving Average* (ARIMA). Model ini memiliki kemampuan prediksi yang sama dengan *Wavelet Neural Network* (WNN), dan *Support Vector Machine* (SVM) dalam memprediksi bencana alam (Zhang et al., 2020). Data kasus Positif COVID-19 dari tanggal 1 September 2020 hingga 3 November 2020 akan digunakan untuk membangun model, dan data dari 4 November hingga 30 November akan digunakan sebagai data uji. Tujuan penelitian ini adalah untuk membantu memprediksi jumlah pasien di kemudian hari dan mempersiapkan sarana kesehatan yang memadai. Serta membantu pemerintah atau pihak yang berwenang dalam mengambil keputusan.

Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data kuantitatif berupa jumlah warga Jakarta per hari yang positif Covid-19 dengan rentang waktu 1 September 2020 hingga 30 November 2020, yaitu sebanyak 91 data. Data dikumpulkan dari halaman *website Open Data Jakarta* (<https://data.jakarta.go.id/dataset?q=Rekap+Data+Covid19+Per+Kelurahan+Provinsi+DKI+Jakarta>). Dataset yang didapat adalah riwayat Covid-19 di Provinsi DKI Jakarta per kelurahan yang berisi berbagai macam variabel dan merupakan penjumlahan dari variabel yang ada seperti : ODP (Proses Pemantauan + Selesai Pemantauan), PDP (Masih Dirawat + Pulang dan Sehat), Positif (Dirawat + Sembuh + Meninggal + *Self Isolation*). Variabel Positif adalah data yang akan dimodelkan dan diramalkan menggunakan model ARIMA (Makridakis et al., 1999b).

Penelitian Terkait

Berikut merupakan beberapa penelitian serupa yang dilakukan untuk meramalkan jumlah kasus COVID-19 di berbagai negara:

Tabel 1
Penelitian terkait

Tahun	Judul	Negara
2020	<i>Forecasting of covid-19 confirmed cases in different countries with arima models</i>	China, Italy, Korea Selatan, Iran, dan Thailand
2020	<i>Arima and nar based prediction model for time series analysis of covid-19 cases in india.</i>	India
2020	<i>Forecasting the Number of Coronavirus (COVID-19) Cases in Ethiopia Using Exponential Smoothing Times Series Model.</i>	Ethiopia

Arima

Auto Regressive Integrated Moving Average atau ARIMA juga dikenal sebagai *Box-Jenkins Model* (Montgomery & Woodall, 2008). ARIMA merupakan salah satu jenis pemodelan yang menjelaskan deret waktu tertentu berdasarkan nilai masa lalunya sendiri (lag nya sendiri) dan kesalahan perkiraan lag, sehingga persamaannya dapat digunakan untuk meramalkan nilai masa depan. Terdapat tiga tahap iterasi dalam melakukan pemodelan ARIMA ini yaitu:

1. Penentuan model tentatif (spesifikasi model) berdasarkan data contoh untuk mengidentifikasi nilai p, d, q .
2. Pendugaan parameter model ARIMA (p,d,q) yang diidentifikasi, yakni penduga nilai ϕ, θ, σ^2_e .

3. Menganalisis hasil untuk melihat kelayakan model.

Notasi model akhirnya berbentuk ARIMA(p,d,q), (Fattah et al., 2018) dengan

1. p adalah orde untuk proses *autoregressive* (AR)
2. d adalah orde yang menyatakan banyaknya proses diferensi yang dilakukan pada data *time series* yang belum stasioner
3. q adalah orde untuk proses *moving average* (MA)

Penentuan parameter pada ARIMA dilakukan dengan melihat fungsi autokorelasi atau disebut dengan *Auto Correlation Function* (ACF) dan fungsi autokorelasi parsial *Partial Correlation Function* (PACF), serta *differencing*. Berikut adalah contoh model ARIMA (1,1,1) terdiri dari: (Makridakis et al., 1999a).

1. Model AR(1)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t$$

$$Y_t - \beta_1 Y_{t-1} = \beta_0 + \epsilon_t$$

Dari persamaan tersebut, jika dilakukan shift mundur (*Backward Shift*) maka :

$$Y_t - \beta_1 B Y_t = \beta_0 + \epsilon_t$$

$$(1 - \beta_1 B) Y_t = \beta_0 + \epsilon_t$$

2. Model MA(1)

$$Y_t = a_0 + e_t - a_1 e_{t-1}$$

dan jika persamaan tersebut dilakukan shift mundur (*Backward Shift*) menjadi:

$$Y_t = a_0 + e_t - a_1 B e_t$$

$$Y_t = a_0 + (1 - a_1 B) e_t.$$

3. *Differencing*

Differencing merupakan proses perbedaan yang bertujuan agar rata-rata pada deret waktu menjadi stasioner. Prosesnya dituliskan sebagai berikut:

$$Y'_t = Y_t - Y_{t-1}$$

$$Y'_t = Y_t - B Y_t$$

$$(1 - B) Y_t$$

Pada model AR(1) dan MA(1) terdapat 2 konstanta yaitu β_0 dan a_0 , sedangkan pada model ARIMA (1,1,1) terdapat sebuah konstanta, anggap μ , maka persamaan bagi ARIMA(1,1,1) adalah:

$$\begin{aligned}
 (1 - B)(1 - \beta_1 B)Y_t &= \mu + e_t - a_1 B e_t \\
 (1 - B)(1 - \beta_1 B)Y_t &= \mu - a_1 B e_t + e_t \\
 (1 - B)(Y_t - \beta_1 B Y_t) &= \mu - a_1 B e_t + e_t \\
 Y_t - \beta_1 B Y_t - B Y_t + \beta_1 B^2 Y_t &= \mu - a_1 B e_t + e_t \\
 Y_t &= \beta_1 B Y_t + B Y_t - \beta_1 B^2 Y_t + \mu - a_1 B e_t + e_t \\
 Y_t &= \beta_1 Y_{t-1} + Y_{t-1} - \beta_1 Y_{t-2} + \mu - a_1 e_{t-1} + e_t
 \end{aligned}$$

Stasioneritas

Pendekatan paling umum untuk membuat deret stasioner adalah dengan mengurangi nilai sebelumnya dari nilai saat ini (Khan & Gupta, 2020). Biasanya dinotasikan dengan *d*. Jika data sudah stasioner, maka *d*=0.

Validasi dan Evaluasi

Model yang dibangun dari data latih (70% data) digunakan untuk memprediksi data uji (30% data). Kriteria Evaluasi yang digunakan adalah *Mean Absolute Error* (MAE) (Yang et al., 2020). Rumus MAE adalah sebagai berikut (Liang et al., 2020)

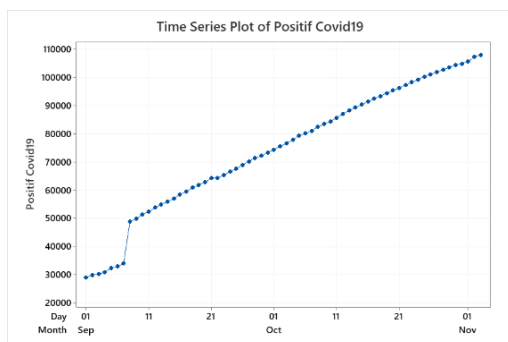
$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}_i|$$

x_i adalah nilai aktual data ke-*i* , *\bar{x}_i* adalah nilai ramalan pada data ke-*i*. *n* adalah jumlah banyaknya data uji. Semakin kecil nilai nya, maka model semakin bagus.

Hasil dan Pembahasan

Data uji memiliki tren naik tanpa indikasi musiman (Grafik 1). Terjadi lonjakan kenaikan jumlah kasus positif COVID-19 di Jakarta pada rentang waktu antara tanggal 1 sampai 11 September 2020 yang mengakibatkan kebutuhan sarana dan prasarana kesehatan ikut meningkat.

Grafik 1. Plot Data Penelitian

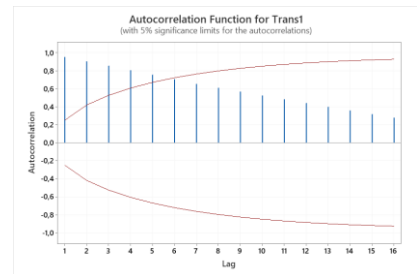


Data awal (Grafik 1) dilakukan transformasi Box-Cox (x-2) satu kali agar stasioner dalam ragam. Data Time Series yang stasioner ditunjukkan oleh nilai Rounded Value (λ) = 1 (Grafik 2). Data hasil transformasi tersebut diperiksa kestasioneran rataannya melalui grafik fungsi autokorelasi (Grafik 3) dan fungsi autokorelasi parsial (Grafik 4). Data stasioner jika semua lag nya berada dalam garis Confident Interval (CI).

Perlakuan differencing dilakukan karena terdapat 5 lag keluar dari garis CI (Grafik 3). Differencing hanya dilakukan 1 kali (d=1) karena data sudah stasioner. Selanjutnya, grafik fungsi autokorelasi (ACF) ini digunakan untuk menentukan parameter AR, dan grafik fungsi autokorelasi parsial digunakan untuk menentukan parameter MA. Berdasarkan hasil analisis, maka didapat beberapa parameter bagi ARIMA, yaitu:

- ARIMA (0,1,1)
- ARIMA (1,1,0)
- ARIMA (1,1,1)

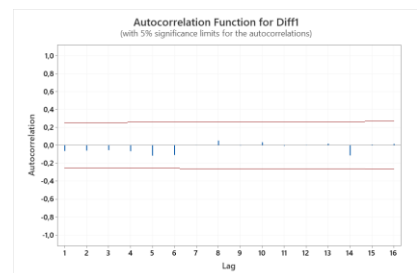
ACF dengan differencing=0



Gambar 1.

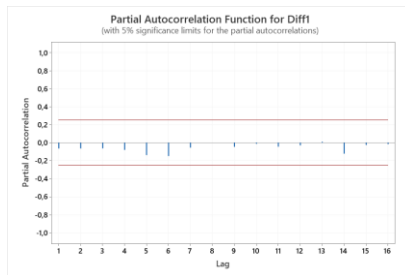
Grafik Analisis ACF dan PACF

ACF dengan differencing=1



Grafik 2. Fungsi ACF sebelum differencing

PACF
dengan
differen
cing=1



Grafik 3. Fungsi ACF dengan $d=1$

Ketiga model ARIMA menghasilkan nilai signifikansi p -value sebagai berikut:

Tabel 2

Nilai p berbagai parameter ARIMA

Model ARIMA	Parameter	Nilai p -value
(1,1,0)	AR(1)	0,938
(0,1,1)	MA(1)	0,937
(1,1,1)	AR(1)	0,000
	MA(1)	0,000

Dapat terlihat dari Tabel 2, bahwa nilai p yang signifikan terdapat pada model ARIMA (1,1,1) karena model AR(1) dan MA(1) memiliki nilai $p < \alpha$ dengan $\alpha=5\%$. Serta, nilai Ljung-Box chi-square statistiknya lebih besar dari $\alpha=5\%$. Dengan demikian, model ARIMA yang akan dipakai untuk peramalan adalah model ARIMA (1,1,1) yang sudah stasioner (error saling bebas) (Chang & Hu, 2019). Peramalan dilakukan untuk 27 hari kedepan, yaitu sebanyak 27 data (30% data *testing*) mulai tanggal 4 – 30 November 2020. Hasil peramalannya ditunjukkan pada Tabel 3 di bawah ini.

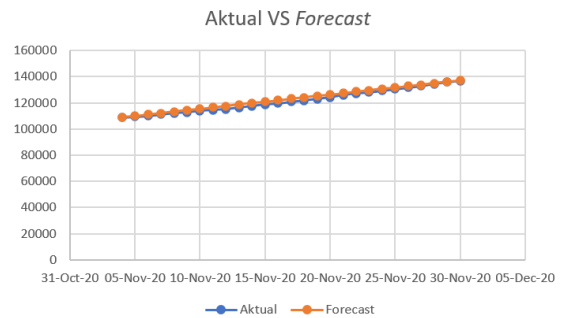
Tabel 3

Nilai *Forecast* selama 27 hari dengan selang kepercayaan 95%.

Period	Forecast	95% Limits		Actual
		Lower	Upper	
4/11/20	108936	105424	112447	108620
5/11/20	110025	105023	115028	109411
6/11/20	111115	104943	117286	110083
7/11/20	112203	105025	119382	111201
8/11/20	113292	105208	121376	112027
9/11/20	114380	105460	123299	112743
10/11/20	115467	105764	125171	113756
11/11/20	116555	106107	127002	114343
12/11/20	117642	106482	128802	115174
13/11/20	118728	106881	130575	116207
14/11/20	119814	107302	132327	117462
15/11/20	120900	107740	134060	118627
16/11/20	121986	108193	135779	119633
17/11/20	123071	108658	137483	120671
18/11/20	124155	109134	139177	121818
19/11/20	125240	109620	140860	123003

20/11/20	126324	110113	142534	124243
21/11/20	127407	110614	144201	125822
22/11/20	128491	111121	145860	127164
23/11/20	129573	111633	147514	128173
24/11/20	130656	112150	149162	129188
25/11/20	131738	112671	150805	130461
26/11/20	132820	113196	152444	131525
27/11/20	133901	113724	154078	132961
28/11/20	134982	114254	155710	134331
29/11/20	136063	114787	157338	135762
30/11/20	137143	115323	158964	136861

Tabel 3 berisi tanggal yang akan diramal jumlah kasus positif COVID-19 nya, lalu variabel *Forecast* merupakan jumlah kasus hasil ramalan, variabel *Actual* merupakan data asli (data uji), sedangkan *upper* dan *lower* adalah batas selang kepercayaan 95% dari model. Berdasarkan perhitungan nilai *Mean Absolute Error* (MAE), didapat hasil 1531. Nilai ini relatif kecil dibanding nilai jumlah kasus positif per hari nya. Perbandingan jumlah warga positif aktual dengan hasil peramalan (*forecast*) digambarkan secara deskriptif melalui Grafik 5. Terlihat bahwa nilai peramalan dan nilai aktual membentuk tren naik, dan titik-titik nilai nya hampir menempel, yang berarti bahwa hasil model ARIMA (1,1,1) mampu meramalkan jumlah kasus Positif COVID-19 pada warga DKI Jakarta pada 4 November hingga 30 November 2020.



Grafik 4. Statistik deskriptif tebaran nilai aktual VS nilai hasil peramalan

Kesimpulan

Model ARIMA (1,1,1) dibangun berdasarkan data harian 1 September hingga 3 November 2020. Model tersebut memiliki nilai MAE yang relatif kecil dan mampu meramalkan jumlah warga DKI Jakarta yang Positif COVID-19 pada 27 hari kedepan dengan sangat baik. Peraturan pemerintah mencanangkan protokol kesehatan pada waktu tersebut adalah hal

yang sangat tepat. Serta masyarakat DKI Jakarta perlu waspada akan wabah ini yang masih juga belum selesai.

BIBLIOGRAFI

- Chang, D.-F., & Hu, H. (2019). Mining gender parity patterns in STEM by using ARIMA model. *ICIC Express Letters, Part B: Applications*, 10(2), 105–112. [Google Scholar](#)
- Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 1847979018808673. [Google Scholar](#)
- Jung, S., Akhmetzhanov, A. R., Hayashi, K., Linton, N. M., Yang, Y., Yuan, B., Kobayashi, T., Kinoshita, R., & Nishiura, H. (2020). Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *Journal of Clinical Medicine*, 9(2), 523. [Google Scholar](#)
- Khan, F. M., & Gupta, R. (2020). Arima and nar based prediction model for time series analysis of covid-19 cases in india. *Journal of Safety Science and Resilience*, 1(1), 12–18. [Google Scholar](#)
- Liang, W., Liang, H., Ou, L., Chen, B., Chen, A., Li, C., Li, Y., Guan, W., Sang, L., & Lu, J. (2020). Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Internal Medicine*, 180(8), 1081–1089. [Google Scholar](#)
- Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1999). Metode dan aplikasi peramalan. *Jakarta: Erlangga*. [Google Scholar](#)
- Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1999). Metode dan Aplikasi Peramalan Jilid 1. *Jakarta: Binarupa Aksara*. [Google Scholar](#)
- Montgomery, D. C., & Woodall, W. H. (2008). An overview of six sigma. *International Statistical Review/Revue Internationale de Statistique*, 329–346. [Google Scholar](#)
- Nemmara, V. V., & Thompson, P. R. (2018). Development of Activity-Based Proteomic Probes for Protein Citrullination. *Activity-Based Protein Profiling*, 233–251. [Google Scholar](#)
- Niehus, R., De Salazar, P. M., Taylor, A. R., & Lipsitch, M. (2020). Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. *MedRxiv*. [Google Scholar](#)
- Yang, Q., Wang, J., Ma, H., & Wang, X. (2020). Research on COVID-19 Based on ARIMA Model—Taking Hubei, China as an example to see the epidemic in Italy. *Journal of Infection and Public Health*, 13(10), 1415–1418. [Google Scholar](#)
- Zhang, Y., Yang, H., Cui, H., & Chen, Q. (2020). Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the Sanjiang Plain, China. *Natural Resources Research*, 29(2), 1447–1464. [Google Scholar](#)

Copyright holder :

Nurul Qomariasih (2021).

First publication right :

Jurnal Syntax Transformation

This article is licensed under:

